Available on CMS information server

CMS CR-2008/059

The Compact Muon Solenoid Experiment

Conference Report

Mailing address: CMS CERN, CH-1211 GENEVA 23, Switzerland

11 August 2008

Measurements of the b-tag performance from data in CMS

Joris Maes

Abstract

Several algorithms are proposed to measure the b-tagging performance from data in CMS. These algorithms can be applied already in a start-up scenario with uncertainties on the estimated b-tag efficiency around 15% for 10 pb^{-1} of data. With about 1000 pb^{-1} of data these uncertainties can be reduced to about 5% depending on our understanding of the detector performance and the physics in proton collisions at 14 TeV.

Presented at *Top 2008 - International Workshop on Top Quark Physics, May 18 - 24, 2008, La Biodola ,Italy*

Measurements of the b-tag performance from data in CMS

J. Maes ([1])(on behalf of the CMS collaboration)

([1]) *Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Brussels, Belgium*

Summary. — Several algorithms are proposed to measure the b-tagging performance from data in CMS. These algorithms can be applied already in a start-up scenario with uncertainties on the estimated b-tag efficiency around 15% for 10 pb^{-1} of data. With about 1000 pb^{-1} of data these uncertainties can be reduced to about 5% depending on our understanding of the detector performance and the physics in proton collisions at 14 TeV.

The CMS detector, one of the multipurpose detectors at the LHC, will collect a divers range of physics processes at the 14 TeV proton-proton collisions. To identify these physics processes, which are predicted by the Standard Model or by theories beyond the Standard Model, jets from b quarks play an important role. To identify jets coming from b quarks the CMS detector is equipped with a silicon based tracking device [1] which is capable to provide an efficient and precise measurement of the trajectories of the charged particles emerging from the collisions. The information of this tracking device and the calorimeters is used in so called b-tagging algorithms to determine the probability of a jet to originate from a b quark. In order not to rely on the performances of these algorithms as predicted from simulation studies, several data-driven methods are proposed to estimate this performance. A divers range of methods exploiting the properties of jets resulting from heavy flavored quarks has been presented to estimate the performance of b-tagging algorithms from collision data detected in CMS.

1. – Measuring the b-tag efficiency from Top Quark events.

The abundantly produced $t\bar{t}$ pairs in proton collisions at the LHC can be used to isolate jet samples with a highly enriched b-jet content, on which the b-jet identification algorithms can be calibrated. Both the semi-leptonic and di-leptonic decay channel are being explored for this method. The event selection requires leptons which are isolated in the tracker and the calorimeter. A minimum number of jets is requested with calibrated $E_T > 25$GeV. A kinematic fit is applied to the event forcing the W boson and top quark mass constraints. To extract a b-enriched jet sample from the selected events in the semi-leptonic channel, only the b-jet coming from the leptonically decaying top quark can be used because one jet was already tagged as a b-jet in the system of the hadronically decaying top quark. Several observables were identified to discriminate between good jet associations and combinatorial background. The likelihood ratio information of all these variables is combined taking into account the correlations between the observables. The jet combination in the event with the highest combined likelihood ratio is chosen and the b-jet is selected as the jet in the leptonic top quark decay. A selection cut on the combined likelihood ratio is used to purify the jet sample in b-quark jets. The optimal cut is chosen in order to minimize the total uncertainty, including for example estimates of the systematic uncertainty due to radiation effects. The b-tag algorithm is applied on the selected jet sample where the purity of the selected jet sample is estimated from simulation. The b-tag efficiency can be measured from this jet sample. A similar analysis path is followed for the di-lepton channels. The information from all decay channels is combined, resulting in expected uncertainties on the estimated b-tag efficiencies shown in figure 1 (left).

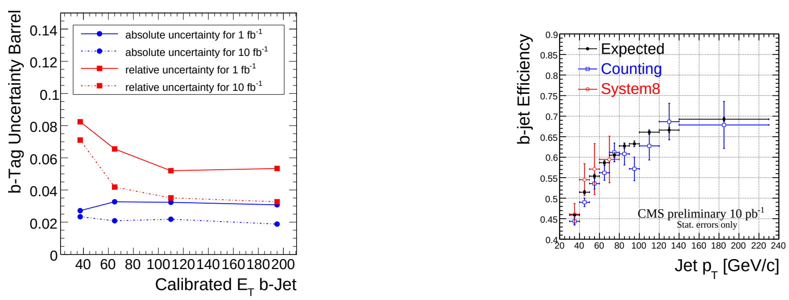

Fig. 1. – Left: Expected combined uncertainty on the estimated b-tag efficiency in the barrel $|\eta| < 1.5$. Right: The b-tagging efficiency as a function of the jet p_T as measured with the Counting and System8 methods compared to the efficiency obtained from the Monte Carlo truth information, statistical errors correspond to an integrated luminosity of about 10 pb^{-1}.

2. – Evaluation of mistags for b-tagging using Negative Tags.

The Track Counting tagger relies on charged particles tracks with a large 3D impact parameter (IP). Impact parameters can be signed as positive (negative) if the associated tracks are produced downstream (upstream) with respect to the primary interaction vertex. Tracks with negative impact parameters can be used to evaluate the tagging efficiency from light (uds) quark and gluon jets, this we define as being the mistagging efficiency [4].

The method is applied on jets with a calibrated $p_T > 20$ GeV from QCD samples. A jet is called taggable if it has at least $n = 1$ associated tracks fulfilling some quality requirements (eg. number of hits, χ^2/ndf, etc.). The taggability is thus simply the ratio between the number of taggable jets and the number of reconstructed jets. For jets of a given flavor, the tagging efficiency is then defined as the number of tagged jets divided by the number of taggable jets.

The mistag efficiency due to light (uds) quark and gluon jets can be evaluated as $\epsilon_{data}^{mistag} = \epsilon_{data}^{-} \cdot R_{light}$, where ϵ_{data}^{-} is the negative tag rate in multi-jet data and $R_{light} = \epsilon_{MC}^{mistag}/\epsilon_{MC}^{-}$ is the ratio between the mistag efficiency of udsg-jets and the negative tag rate of all (udsg+c+b) jets in the simulation. The c and b fractions can be significantly reduced by applying a positive tag veto: the current negative tag jet is rejected if it has any track with IP/$\sigma_{IP} > 4$. With these numbers we can obtain the mistag rates on data as $\epsilon_{data}^{mistag}(p_T, \eta) = (\epsilon_{MC}^{mistag}(p_T, \eta) \cdot \epsilon_{data}^{-})/\epsilon_{MC}^{-}$. For the *medium* working point the total systematic uncertainty on the udsg-mistag rate is estimated to be 8.8%, 7.6% and 5.9% for respectively 10, 100 and 1000 pb^{-1}.

3. – Measuring b-tagging performance using Jets containing Muons.

The analyses are based on samples that have at least two reconstructed jets and a non-isolated muon close to one of the jets with $\Delta R(\mu, jet) < 0.4$. The variable p_{Trel} is defined as the transverse momentum of the muon relative to the direction of the total muon-jet momentum vector.

- The p_{Trel} *method* relies directly on a fit to the p_{Trel} distribution of the muon before and after tagging the muon-jet. Templates were obtained for different ranges of jet p_T and $|\eta|$. The p_{Trel} distribution of the muons is fitted with a linear combination of the b and c+light jet templates. The process is repeated after tagging the muon-jet. The b-tagging efficiency is calculated as the ratio between the number of b jets after and before tagging, as determined by the p_{Trel} fits [3].

- The *Counting Method* also relies on the p_{Trel} observable fits but uses additional information derived from data. It assumes that the away-jets in the sample are dominated by light jets, and that the average probability of tagging them can be estimated from light jets data sample with negative impact parameter with respect to the interaction point [3].

- The third method, so-called *System8 Method* does not rely on the p_{Trel} fits to extract the b jet content of the samples. It consists on solving a system of eight equations constructed from the total number of events in two samples with different b jet content, before and after tagging with two b-tagging algorithms [3].

The efficiencies are measured as a function of the jet p_T and $|\eta|$ and agree rather well with each other and with the efficiency obtained from the MC truth. This is illustrated in figure 1 (right). The total uncertainty on the estimated b-tagging efficiency are about equal for all three methods and are equal to 15%, 10% and 5% for respectively 10, 100 and 1000 pb^{-1}. The *System8 Method* performance is slightly less sensitive to systematic uncertainties at start-up for 10 pb^{-1} as it depends less on Monte Carlo simulation, however it produces results in a limited range of p_T.

REFERENCES

[1] CMS Physics: Technical Design Report, Vol. 1, CERN-LHCC-2006-001/021 (2006).

[2] S. Lowette, J. D'Hondt, J. Heyninck and P. Vanlaer, "Offline calibration of b-jet identification efficiencies", CERN-CMS-NOTE-2006-013.

[3] The CMS Collaboration, Performance Measurement of b-tagging Algorithms Using Data containing Muons within Jets, http://cms-physics.web.cern.ch/cms-physics/public/BTV-07-001-pas.pdf

[4] The CMS Collaboration, Evaluation of udsg Mistags for b-tagging using Negative Tags, http://cms-physics.web.cern.ch/cms-physics/public/BTV-07-002-pas.pdf